

An iterative algorithm for sparse and constrained recovery with applications to divergence-free current reconstructions in magneto-encephalography

Ignace Loris and Caroline Verhoeven

February 16, 2012

Abstract

We propose an iterative algorithm for the minimization of a ℓ_1 -norm penalized least squares functional, under additional linear constraints. The algorithm is fully explicit: it uses only matrix multiplications with the three matrices present in the problem (in the linear constraint, in the data misfit part and in penalty term of the functional). None of the three matrices must be invertible. Convergence is proven in a finite-dimensional setting. We apply the algorithm to a synthetic problem in magneto-encephalography where it is used for the reconstruction of divergence-free current densities subject to a sparsity promoting penalty on the wavelet coefficients of the current densities. We discuss the effects of imposing zero divergence and of imposing joint sparsity (of the vector components of the current density) on the current density reconstruction.

1 Introduction

In magneto-encephalography (MEG) an image of an electrical current density is reconstructed from measurements of the magnetic field outside the scalp. The magnetic field generated by these currents is very weak compared to the environment; special precautions are taken to minimize these external effects on the observed data. Another characteristic of MEG imaging is the very low number of data: typically only a few hundred of measurements are taken. When using a current density representation of reasonable size (spatial resolution) these data are not complete. This implies having to

solve an underdetermined system of equations. Extra conditions need to be imposed to define a unique current density reconstruction.

In [12] the use of a sparsity promoting penalty, together with an efficient representation of the current density in terms of wavelets [9, 18] was proposed. In other words, the assumption that the unknown current density can be represented with a small number of non-zero wavelet coefficients, was used as a priori information to regularize the inversion.

The regularization of the MEG inverse problem by a sparsity assumption was carried out in practice in [12] by adding an ℓ_1 -norm penalty term to a quadratic cost function for the data misfit. The ℓ_1 -norm is a popular sparsity promoting penalty [5, 4] as it allows for convex optimization techniques to be used instead of algorithms with combinatorial complexity. It was shown in [10] that such a penalty regularizes the linear inverse problem, and convergence of an iterative soft-thresholding algorithm for the minimization of an ℓ_1 -norm penalized least squares functional was proven in a Hilbert space setting. In the present work we extend the method of [12, 10] to incorporate linear constraints (e.g. to impose zero divergence on the reconstructed current densities) and to handle a more general sparsity promoting penalty.

In the first, mathematical, part of this paper we propose a new iterative algorithm for the minimization of an ℓ_1 -norm penalized least squares functional, under additional linear constraints:

$$\hat{x} = \arg \min_{Bx=b} \|Kx - y\|^2 + 2\lambda \|Ax\|_1. \quad (1)$$

Here K is the matrix that defines the linear relation between the unknown model x and the data y ; $Bx = b$ is a linear constraint on the solution. A is a matrix mixing the variables in the non-smooth penalty term $\|Ax\|_1$. It need not be invertible in our approach (e.g. $A = \text{grad}$ would correspond to a total variation penalty [20]). We write the variational equations corresponding to this problem and derive a simple iterative algorithm. This algorithm consists of a single loop and each step in the loop is given explicitly in terms of matrix multiplications by K , A and B (and their transposes). We prove the convergence of this algorithm for general K , A and B (subject to a bound on their norms) in a finite-dimensional setting. In problem (1) the ℓ_1 -norm penalty may be replaced by another convex function $H(Ax)$. The proposed algorithm can be modified to apply to this case as well, as long the proximity operator of H is known.

The proposed algorithm reduces to the generalized iterative soft-thresholding algorithm of [17] when the linear constraints $Bx = b$ are removed. We will indicate below in which special cases (e.g. $A = \text{Id}$) our algorithm is also derivable by the method in [23]. In these special cases we also comment on

the conditions on the matrices that are necessary for guaranteeing convergence; in particular, we indicate where our conditions on the matrices K , A or B are less strict than the ones derivable from the work in [23].

In the second part of this paper, we apply the algorithm to an inverse problem loosely based on magneto-encephalography. We shall assume a linear relationship between an unknown current density \vec{J} and a measured magnetic field \vec{B} in the form of the Biot-Savart law. Furthermore we assume that the data are contaminated by Gaussian noise. As in [12] we will impose sparsity on the wavelet expansion of the current density \vec{J} . That is, we will use the proposed iterative algorithm to solve problem (1) where x is the current density \vec{J} , K a matrix corresponding to a discretized Biot-Savart law, A the wavelet transform, and $Bx = b$ represents the linear constraints $\text{div}(\vec{J}) = 0$. In this last point lies the main difference with the simulations in [12]: we shall incorporate into the reconstruction procedure the assumption that the current density is divergence-free; this was not done in [12]. Our approach does not use divergence-free wavelets [14, 15, 22], but relies on an explicit linear constraint for finding a divergence-free reconstruction.

On a synthetic problem, we investigate the effect of the divergence-free nature of the current distribution on the sparse reconstruction. That is, we compare reconstructions from the same measurement data with and without the constraint. Secondly, we investigate the effect on the reconstruction of imposing a “joint sparsity” [13] condition on the two vector components of the current density. This means that at each position both vector components are simultaneously zero or simultaneously non-zero. Joint sparsity in MEG was first discussed in [12] as a way of improving reconstruction quality. The proposed iterative algorithm can also handle penalties that promote joint sparsity (see discussion at the end of Section 3).

Besides MEG, other applications of ℓ_1 -penalized least squares under linear constraints exist. One is found in the portfolio selection problem described in [3]. Such problems are often of a smaller size (fewer variables) and can sometimes also be solved via a non-iterative procedure. Another application of the minimization problem (1) is found in the formulation of a modified Total Variation model in the context of image processing tasks [16].

In this paper we will make frequent use of the (non-linear) soft-thresholding operator which is defined by:

$$S_\lambda(z) = \begin{cases} z - \frac{z}{|z|}\lambda & |z| > \lambda \\ 0 & |z| \leq \lambda \end{cases} \quad (2)$$

and of the projection on the ℓ_∞ ball of radius λ defined by:

$$P_\lambda(z) = \begin{cases} \frac{z}{|z|}\lambda & |z| > \lambda \\ z & |z| \leq \lambda. \end{cases} \quad (3)$$

We have that:

$$S_\lambda(z) + P_\lambda(z) = z \quad (4)$$

for all z . We set $B_\lambda^\infty = \{u \text{ with } \|u\|_\infty \leq \lambda\}$ (ℓ_∞ -ball of radius λ).

2 Description of the iterative algorithm

The variational equations of the minimization problem (1) can be obtained by the introduction of Lagrange multipliers v :

$$\min_x \|Kx - y\|^2 + 2\lambda\|Ax\|_1 - 2\langle v, Bx - b \rangle \quad \text{and} \quad Bx = b. \quad (5)$$

Derivation with respect to x yields:

$$K^T(Kx - y) + A^T w - B^T v = 0 \quad \text{and} \quad Bx = b,$$

where w is an element of the subdifferential of $\lambda\|Ax\|_1$, i.e. $w_i = \lambda(Ax)_i/|(Ax)_i|$ if $(Ax)_i \neq 0$ and $|w_i| \leq \lambda$ if $(Ax)_i = 0$. This can be written more compactly as $(Ax)_i = S_\lambda(w_i + (Ax)_i)$ or equivalently $w_i = P_\lambda(w_i + (Ax)_i)$. We find that the variational equations corresponding to the problem (1) therefore are:

$$K^T(Kx - y) + A^T w - B^T v = 0, \quad w = \mathbb{P}_\lambda(w + Ax) \quad \text{and} \quad Bx = b, \quad (6)$$

where $\mathbb{P}_\lambda(u)$ corresponds to the application of P_λ (defined in (formula 3)) on each component of u . We assume that a solution to these equations exists and try to derive an iterative algorithm that converges to such a solution. By writing $\lambda\|Ax\|_1 = \max_{w \in B_\lambda^\infty} \langle w, Ax \rangle$, the minimization problem (1) can be expressed as:

$$\min_{x, Bx=b} \max_{w \in B_\lambda^\infty} F(x, w, v), \quad (7)$$

where we have set:

$$F(x, w, v) = \|Kx - y\|^2 + 2\langle w, Ax \rangle - 2\langle v, Bx - b \rangle. \quad (8)$$

We write the variational equations (6) as fixed-point equations:

$$\begin{cases} w &= \mathbb{P}_\lambda(w + Ax) \\ x &= x + K^T(y - Kx) + B^T v - A^T w \\ v &= v - \frac{1}{\alpha}(Bx - b) \end{cases} \quad (9)$$

and study the predictor-corrector scheme:

$$\begin{cases} \bar{v}^{n+1} &= v^n - (Bx^n - b) \\ \bar{x}^{n+1} &= x^n + K^T(y - Kx^n) + B^T\bar{v}^{n+1} - A^Tw^n \\ w^{n+1} &= \mathbb{P}_\lambda(w^n + A\bar{x}^{n+1}) \\ x^{n+1} &= x^n + K^T(y - Kx^n) + B^T\bar{v}^{n+1} - A^Tw^{n+1} \\ v^{n+1} &= v^n - \frac{1}{\alpha}(Bx^{n+1} - b). \end{cases} \quad (10)$$

There is a predictor-corrector step on the variables v and x but not on w . Clearly, the fixed-point of this iteration is a solution to the variational equations (6). Moreover the algorithm is fully explicit: Each step only requires the application of the matrices K, A, B (and their transposes) and a simple projection \mathbb{P}_λ (see formula (3)). There is no non-trivial sub-problem to solve in each step (such as e.g. solving a linear system of equations). In other words, there is no inner loop required for any of the lines in (10). In the next section we show that, under certain conditions on the operators A, B and K , and on the parameter α , the proposed algorithm (10) converges to a solution of the variational equations (6), and to a minimizer of the functional (1).

For the special case when $B = 0$ and $b = 0$ (absence of linear constraints), the algorithm (10) reduces to:

$$\begin{cases} \bar{x}^{n+1} &= x^n + K^T(y - Kx^n) - A^Tw^n \\ w^{n+1} &= \mathbb{P}_\lambda(w^n + A\bar{x}^{n+1}) \\ x^{n+1} &= x^n + K^T(y - Kx^n) - A^Tw^{n+1}. \end{cases} \quad (11)$$

This algorithm was presented in [17] to solve the problem:

$$\hat{x} = \arg \min_x \|Kx - y\|^2 + 2\lambda\|Ax\|_1. \quad (12)$$

An important application is the total variation penalty in image analysis ($A = \text{grad}$). A similar algorithm (with predictor-corrector step on the w variable) was proposed independently in [2] for Poisson data. When $A = \text{Id}$, and using relation (4), algorithm (11) further simplifies to the traditional iterative soft-thresholding algorithm:

$$x^{n+1} = \mathbb{S}_\lambda(x^n + K^T(y - Kx^n)), \quad (13)$$

where $\mathbb{S}_\lambda(u)$ corresponds to the application of S_λ (defined in formula (2)) on each component of u . This algorithm was discussed in [10] for solving

$$\min_x \|Kx - y\|^2 + 2\lambda\|x\|_1. \quad (14)$$

An accelerated version of this algorithm, the Fast Iterative Soft-Thresholding Algorithm (FISTA), was derived in [1]. Many other algorithms exist as well.

On the other hand, when the constraints $Bx = b$ are maintained, and with A equal to the identity, the algorithm (10) reduces to a constrained version of the iterative soft-thresholding algorithm:

$$\begin{cases} \bar{v}^{n+1} = v^n - (Bx^n - b) \\ x^{n+1} = \mathbb{S}_\lambda(x^n + K^T(y - Kx^n) + B^T\bar{v}^{n+1}) \\ v^{n+1} = v^n - \frac{1}{\alpha}(Bx^{n+1} - b), \end{cases} \quad (15)$$

for the problem

$$\hat{x} = \arg \min_{Bx=b} \|Kx - y\|^2 + 2\lambda\|x\|_1. \quad (16)$$

We will use algorithm (15) for an application in magneto-encephalography in section 5. Although it was not included in [23], algorithm (15) for problem (16), could also have been derived from the Bregman framework of [23]. At the end of section 3 we will comment on the difference between conditions of convergence that the matrices K and B have to satisfy to guarantee convergence of (15) in our approach and in [23].

Finally, by setting $K = 0$ and $A = \text{Id}$ in problem (1) one recovers the so-called ℓ_1 basis pursuit problem [6]:

$$\arg \min_{Bx=b} \|x\|_1 \quad (17)$$

for which the algorithms (10) and (15) reduce to:

$$\begin{cases} \bar{v}^{n+1} = v^n - (Bx^n - b) \\ x^{n+1} = \mathbb{S}_\lambda(x^n + B^T\bar{v}^{n+1}) \\ v^{n+1} = v^n - \frac{1}{\alpha}(Bx^{n+1} - b), \end{cases} \quad (18)$$

This algorithm was discussed in [23] (using different notation and auxiliary variables) in a Bregman framework.

Taking these special cases into account we can say that the proposed algorithm (10) combines the generalized iterative soft-thresholding algorithm (11) of [17] with the basis pursuit algorithm (18) into a single unified algorithm.

3 Proof of convergence

In this section, we prove the convergence of algorithm (10) and show that this yields a minimum of functional (1).

Lemma 1. *If $u^+ = \mathbb{P}_\lambda(u^- + \Delta)$, with \mathbb{P}_λ the projection on the convex set B_λ^∞ , then*

$$\|u - u^+\|^2 \leq \|u - u^-\|^2 - \|u^- - u^+\|^2 - 2\langle u - u^+, \Delta \rangle \quad (19)$$

for all $u \in B_\lambda^\infty$.

Proof. As \mathbb{P}_λ is the projection on a non-empty closed convex set one has:

$$\langle u - \mathbb{P}_\lambda(u'), u' - \mathbb{P}_\lambda(u') \rangle \leq 0$$

for all $u \in B_\lambda^\infty$ and all u' . Choosing $u' = u^- + \Delta$ and $\mathbb{P}_\lambda(u') = u^+$ yields

$$\langle u - u^+, u^- + \Delta - u^+ \rangle \leq 0.$$

Replacing $\langle u - u^+, u^- - u^+ \rangle$ by $(\|u - u^+\|^2 + \|u^- - u^+\|^2 - \|u - u^-\|^2) / 2$ yields

$$\|u - u^+\|^2 + \|u^- - u^+\|^2 - \|u - u^-\|^2 + 2\langle u - u^+, \Delta \rangle \leq 0$$

which is the desired result. \square

The operator \mathbb{P}_λ and the convex set B_λ^∞ in Lemma 1 may be replaced with a projection on any non-empty closed convex set. In particular, when $u^+ = u^- + \Delta$ (i.e. B_λ^∞ replaced by the whole space and \mathbb{P}_λ replaced by the identity), one has that:

$$\|u - u^+\|^2 = \|u - u^-\|^2 - \|u^- - u^+\|^2 - 2\langle u - u^+, \Delta \rangle \quad (20)$$

for all u .

Lemma 2. *If $(x^{n+1}, w^{n+1}, v^{n+1})$ and (x^n, w^n, v^n) are related by iteration (10) then*

$$\begin{aligned} \|x - x^{n+1}\|^2 + \|w - w^{n+1}\|^2 + \alpha\|v - v^{n+1}\|^2 \leq & \\ & \|x - x^n\|^2 + \|w - w^n\|^2 + \alpha\|v - v^n\|^2 - \|x^n - x^{n+1}\|^2 \\ & - \|w^n - w^{n+1}\|^2 - \alpha\|v^n - v^{n+1}\|^2 - \|K(x - x^n)\|^2 \\ & + \|K(x^n - x^{n+1})\|^2 - \|B(x - x^n)\|^2 + \|B(x - x^{n+1})\|^2 \\ & + \|B(x^n - x^{n+1})\|^2 - \|A^T(w - w^n)\|^2 \\ & + \|A^T(w - w^{n+1})\|^2 + \|A^T(w^{n+1} - w^n)\|^2 \\ & + F(x, w^{n+1}, v^{n+1}) - F(x^{n+1}, w, v) \\ & + 2\frac{\alpha-1}{\alpha}\langle B(x - x^{n+1}), Bx^{n+1} - b \rangle \end{aligned} \quad (21)$$

for all x, v and all $w \in B_\lambda^\infty$.

Proof. From Lemma 1 and equation (20), we find:

$$\begin{aligned} \|x - x^{n+1}\|^2 + \|w - w^{n+1}\|^2 + \alpha\|v - v^{n+1}\|^2 \leq & \|x - x^n\|^2 + \|w - w^n\|^2 \\ & + \alpha\|v - v^n\|^2 - \|x^n - x^{n+1}\|^2 - \|w^n - w^{n+1}\|^2 - \alpha\|v^n - v^{n+1}\|^2 \\ & - 2\langle x - x^{n+1}, K^T(y - Kx^n) \rangle - 2\langle x - x^{n+1}, B^T \bar{v}^{n+1} - A^T w^{n+1} \rangle \\ & - 2\langle w - w^{n+1}, A \bar{x}^{n+1} \rangle + 2\langle v - v^{n+1}, Bx^{n+1} - b \rangle. \end{aligned}$$

As (10) implies that $\bar{v}^{n+1} = v^{n+1} + B(x^{n+1} - x^n) + \frac{1-\alpha}{\alpha}(Bx^{n+1} - b)$ and that $\bar{x}^{n+1} = x^{n+1} + A^T(w^{n+1} - w^n)$, this can be written as:

$$\begin{aligned} & \|x - x^{n+1}\|^2 + \|w - w^{n+1}\|^2 + \alpha\|v - v^{n+1}\|^2 \leq \|x - x^n\|^2 + \|w - w^n\|^2 \\ & + \alpha\|v - v^n\|^2 - \|x^n - x^{n+1}\|^2 - \|w^n - w^{n+1}\|^2 - \alpha\|v^n - v^{n+1}\|^2 \\ & - 2\langle K(x - x^{n+1}), y - Kx^n \rangle + 2\langle A(x - x^{n+1}), w^{n+1} \rangle \\ & - 2\langle B(x - x^{n+1}), v^{n+1} \rangle - 2\langle B(x - x^{n+1}), B(x^{n+1} - x^n) \rangle \\ & + 2\frac{\alpha-1}{\alpha}\langle B(x - x^{n+1}), Bx^{n+1} - b \rangle - 2\langle w - w^{n+1}, Ax^{n+1} \rangle \\ & - 2\langle A^T(w - w^{n+1}), A^T(w^{n+1} - w^n) \rangle + 2\langle v - v^{n+1}, Bx^{n+1} - b \rangle. \end{aligned}$$

The terms in $\langle v^{n+1}, Bx^{n+1} \rangle$ and $\langle w^{n+1}, Ax^{n+1} \rangle$ drop and the remaining terms can be re-arranged to yield:

$$\begin{aligned} & \|x - x^{n+1}\|^2 + \|w - w^{n+1}\|^2 + \alpha\|v - v^{n+1}\|^2 \leq \|x - x^n\|^2 + \|w - w^n\|^2 \\ & + \alpha\|v - v^n\|^2 - \|x^n - x^{n+1}\|^2 - \|w^n - w^{n+1}\|^2 - \alpha\|v^n - v^{n+1}\|^2 \\ & - 2\langle K(x - x^{n+1}), y - Kx^n \rangle - 2\langle A^T(w - w^{n+1}), A^T(w^{n+1} - w^n) \rangle \\ & - 2\langle B(x - x^{n+1}), B(x^{n+1} - x^n) \rangle + 2\frac{\alpha-1}{\alpha}\langle B(x - x^{n+1}), Bx^{n+1} - b \rangle \\ & + 2\langle w^{n+1}, Ax \rangle - 2\langle w, Ax^{n+1} \rangle - 2\langle v^{n+1}, Bx - b \rangle + 2\langle v, Bx^{n+1} - b \rangle. \end{aligned}$$

By rewriting the following inner products:

$$\begin{aligned} -2\langle K(x - x^{n+1}), y - Kx^n \rangle &= \|Kx - y\|^2 - \|Kx^{n+1} - y\|^2 \\ &\quad - \|K(x - x^n)\|^2 + \|K(x^n - x^{n+1})\|^2 \\ -2\langle A^T(w - w^{n+1}), A^T(w^{n+1} - w^n) \rangle &= \|A^T(w - w^{n+1})\|^2 \\ &\quad + \|A^T(w^{n+1} - w^n)\|^2 - \|A^T(w - w^n)\|^2 \\ -2\langle B(x - x^{n+1}), B(x^{n+1} - x^n) \rangle &= \|B(x - x^{n+1})\|^2 \\ &\quad + \|B(x^{n+1} - x^n)\|^2 - \|B(x - x^n)\|^2, \end{aligned}$$

and by using the expression (8) of $F(x, v, w)$ in:

$$\begin{aligned} 2\langle w^{n+1}, Ax \rangle - 2\langle w, Ax^{n+1} \rangle - 2\langle v^{n+1}, Bx - b \rangle + 2\langle v, Bx^{n+1} - b \rangle = \\ F(x, w^{n+1}, v^{n+1}) - F(x^{n+1}, w, v) - \|Kx - y\|^2 + \|Kx^{n+1} - y\|^2, \end{aligned}$$

the previous inequality can be written as (21), which proves the lemma. \square

Lemma 3. *If $(\hat{x}, \hat{w}, \hat{v})$ satisfies the variational equations (6), then*

$$F(\hat{x}, w, v) - F(x, \hat{w}, \hat{v}) \leq -\|K(x - \hat{x})\|^2 \quad (22)$$

for all x, v and all $w \in B_\lambda^\infty$.

Proof. To prove inequality (22), we use Lemma 2, where we replace both (x^n, w^n, v^n) and $(x^{n+1}, w^{n+1}, v^{n+1})$ by $(\hat{x}, \hat{w}, \hat{v})$; this is allowed because $(\hat{x}, \hat{w}, \hat{v})$

satisfies (9) which are the fixed-point equations of algorithm (10). Then relation (21) becomes:

$$\begin{aligned} \|x - \hat{x}\|^2 + \|w - \hat{w}\|^2 + \alpha\|v - \hat{v}\|^2 &\leq \|x - \hat{x}\|^2 + \|w - \hat{w}\|^2 + \alpha\|v - \hat{v}\|^2 \\ &\quad - \|K(x - \hat{x})\|^2 \\ &\quad - \|B(x - \hat{x})\|^2 + \|B(x - \hat{x})\|^2 \\ &\quad - \|A^T(w - \hat{w})\|^2 + \|A^T(w - \hat{w})\|^2 \\ &\quad + F(x, \hat{w}, \hat{v}) - F(\hat{x}, w, v), \end{aligned} \quad (23)$$

for all x, v and all $w \in B_\lambda^\infty$. This implies inequality (22). \square

We now show that a solution of equations (6) solves the minimization problem (1).

Theorem 1. *If $(\hat{x}, \hat{w}, \hat{v})$ satisfies the variational equations (6) then \hat{x} is a solution of the minimization problem (1).*

Proof. If $(\hat{x}, \hat{w}, \hat{v})$ is a solution of (6) it follows from Lemma 3 that $F(\hat{x}, w, v) \leq F(x, \hat{w}, \hat{v})$ for all x, v and all $w \in B_\lambda^\infty$, which means:

$$\|K\hat{x} - y\|^2 + 2\langle w, A\hat{x} \rangle \leq \|Kx - y\|^2 + 2\langle \hat{w}, Ax \rangle - 2\langle \hat{v}, Bx - b \rangle.$$

Taking the maximum over $w \in B_\lambda^\infty$ in the left hand side gives

$$\|K\hat{x} - y\|^2 + 2\lambda\|A\hat{x}\|_1 \leq \|Kx - y\|^2 + 2\langle \hat{w}, Ax \rangle - 2\langle \hat{v}, Bx - b \rangle$$

and since $\langle \hat{w}, Ax \rangle \leq \max_{\|\tilde{w}\|_\infty \leq \lambda} \langle \tilde{w}, Ax \rangle = \|Ax\|_1$ one finds:

$$\|K\hat{x} - y\|^2 + 2\lambda\|A\hat{x}\|_1 \leq \|Kx - y\|^2 + 2\lambda\|Ax\|_1 - 2\langle \hat{v}, Bx - b \rangle.$$

for all x, v . As we minimize under the condition that $Bx = b$, we have that $\langle \hat{v}, Bx - b \rangle = 0$ and find

$$\|K\hat{x} - y\|^2 + 2\lambda\|A\hat{x}\|_1 \leq \|Kx - y\|^2 + 2\lambda\|Ax\|_1,$$

for all x for which $Bx = b$. As $B\hat{x} = b$ this proves the theorem. \square

Theorem 2. *If the set $\{x, \text{ with } Bx = b\}$ is non-empty, $\|AA^T\| < 1$, $\|\frac{1}{2}K^TK + B^TB\| < 1$ and $\alpha > \frac{1}{2}$, then the sequence $(x^n, w^n, v^n)_{n \in \mathbb{N}}$ defined by the iteration*

$$\begin{cases} \bar{v}^{n+1} &= v^n - (Bx^n - b) \\ \bar{x}^{n+1} &= x^n + K^T(y - Kx^n) + B^T\bar{v}^{n+1} - A^Tw^n \\ w^{n+1} &= \mathbb{P}_\lambda(w^n + A\bar{x}^{n+1}) \\ x^{n+1} &= x^n + K^T(y - Kx^n) + B^T\bar{v}^{n+1} - A^Tw^{n+1} \\ v^{n+1} &= v^n - \frac{1}{\alpha}(Bx^{n+1} - b) \end{cases} \quad (24)$$

converges to a solution $(x^\dagger, w^\dagger, v^\dagger)$ of the variational equations (6), and a solution of the minimization problem (1).

Proof. If $\{x, \text{ with } Bx = b\}$ is not empty, there is a solution to (1), implying that there exists a solution $(\hat{x}, \hat{w}, \hat{v})$ to the variational equations (9). We now use Lemma 2 with $(x, w, v) = (\hat{x}, \hat{w}, \hat{v})$ and find:

$$\begin{aligned} & \|\hat{x} - x^{n+1}\|^2 + \|\hat{w} - w^{n+1}\|^2 + \alpha\|\hat{v} - v^{n+1}\|^2 \leq \|\hat{x} - x^n\|^2 + \|\hat{w} - w^n\|^2 \\ & + \alpha\|\hat{v} - v^n\|^2 - \|x^n - x^{n+1}\|^2 - \|w^n - w^{n+1}\|^2 - \alpha\|v^{n+1} - v^n\|^2 \\ & - \|K(\hat{x} - x^n)\|^2 + \|K(x^n - x^{n+1})\|^2 - \|B(\hat{x} - x^n)\|^2 + \|B(\hat{x} - x^{n+1})\|^2 \\ & + \|B(x^n - x^{n+1})\|^2 - \|A^T(\hat{w} - w^n)\|^2 + \|A^T(\hat{w} - w^{n+1})\|^2 \\ & + \|A^T(w^{n+1} - w^n)\|^2 - \|K(\hat{x} - x^{n+1})\|^2 + 2\frac{1-\alpha}{\alpha}\langle B(x^{n+1} - \hat{x}), Bx^{n+1} - b \rangle, \end{aligned}$$

where we also used relation (22) with $(x, w, v) = (x^{n+1}, w^{n+1}, v^{n+1})$. As $B\hat{x} = b$ we now use that:

$$2\frac{1-\alpha}{\alpha}\langle B(x^{n+1} - \hat{x}), Bx^{n+1} - b \rangle = 2\frac{1-\alpha}{\alpha}\|Bx^{n+1} - b\|^2 = 2\alpha(1-\alpha)\|v^{n+1} - v^n\|^2,$$

and:

$$\begin{aligned} -\|K(\hat{x} - x^n)\|^2 - \|K(x^{n+1} - \hat{x})\|^2 &= -\frac{1}{2}\|K(x^{n+1} - x^n)\|^2 \\ &\quad - \frac{1}{2}\|K(2\hat{x} - x^{n+1} - x^n)\|^2 \\ &\leq -\frac{1}{2}\|K(x^{n+1} - x^n)\|^2, \end{aligned}$$

and reorder to obtain:

$$\begin{aligned} & \|\hat{x} - x^{n+1}\|^2 - \|B(\hat{x} - x^{n+1})\|^2 + \|\hat{w} - w^{n+1}\|^2 - \|A^T(\hat{w} - w^{n+1})\|^2 \\ & + \alpha\|\hat{v} - v^{n+1}\|^2 \leq \|\hat{x} - x^n\|^2 - \|B(\hat{x} - x^n)\|^2 + \|\hat{w} - w^n\|^2 \\ & - \|A^T(\hat{w} - w^n)\|^2 + \alpha\|\hat{v} - v^n\|^2 \\ & - \left(\|x^{n+1} - x^n\|^2 - \frac{1}{2}\|K(x^{n+1} - x^n)\|^2 - \|B(x^n - x^{n+1})\|^2 \right) \\ & - \left(\|w^{n+1} - w^n\|^2 - \|A^T(w^{n+1} - w^n)\|^2 \right) \\ & - \alpha(2\alpha - 1)\|v^{n+1} - v^n\|^2. \end{aligned}$$

As we assume that $\|AA^T\| < 1$ and $\|\frac{1}{2}K^TK + B^TB\| < 1$ we can introduce regular square matrices L , U and V by $L^TL = 1 - \frac{1}{2}K^TK - B^TB$, $U^TU = 1 - B^TB$ and $V^TV = 1 - AA^T$ to find:

$$\begin{aligned} & \|U(\hat{x} - x^{n+1})\|^2 + \|V(\hat{w} - w^{n+1})\|^2 + \alpha\|\hat{v} - v^{n+1}\|^2 \\ & \leq \|U(\hat{x} - x^n)\|^2 + \|V(\hat{w} - w^n)\|^2 + \alpha\|\hat{v} - v^n\|^2 \\ & \quad - \|L(x^n - x^{n+1})\|^2 - \|V(w^n - w^{n+1})\|^2 \\ & \quad - \alpha(2\alpha - 1)\|v^{n+1} - v^n\|^2. \end{aligned} \tag{25}$$

Summing from M to $N > M$ one finds:

$$\begin{aligned}
& \|U(\hat{x} - x^{N+1})\|^2 + \|V(\hat{w} - w^{N+1})\|^2 + \alpha\|\hat{v} - v^{N+1}\|^2 \\
& \leq \|U(\hat{x} - x^M)\|^2 + \|V(\hat{w} - w^M)\|^2 + \alpha\|\hat{v} - v^M\|^2 \\
& \quad - \sum_{n=M}^N \left(\|L(x^n - x^{n+1})\|^2 + \|V(w^n - w^{n+1})\|^2 \right. \\
& \quad \left. + \alpha(2\alpha - 1)\|v^{n+1} - v^n\|^2 \right). \tag{26}
\end{aligned}$$

Since $\alpha > \frac{1}{2}$ the summation on the right hand side is negative. As U and V are invertible, it follows that the sequence (x^n, w^n, v^n) is bounded. And, as we work in a finite dimensional space, there is a convergent subsequence $(x_{n_j}, w_{n_j}, v_{n_j}) \xrightarrow{j \rightarrow \infty} (x^\dagger, w^\dagger, v^\dagger)$. It also follows from inequality (26) that:

$$\begin{aligned}
& \sum_{n=M}^N \left(\|L(x^n - x^{n+1})\|^2 + \|V(w^n - w^{n+1})\|^2 + \alpha(2\alpha - 1)\|v^{n+1} - v^n\|^2 \right) \\
& \leq \|U(\hat{x} - x^M)\|^2 + \|V(\hat{w} - w^M)\|^2 + \alpha\|\hat{v} - v^M\|^2
\end{aligned}$$

As $\alpha > \frac{1}{2}$, $\|L(x^n - x^{n+1})\|^2$, $\|V(w^n - w^{n+1})\|^2$ and $\|v^{n+1} - v^n\|^2$ tend to zero for large n , which implies that $\|x^n - x^{n+1}\|^2$ and $\|w^n - w^{n+1}\|^2$ tend to zero as well. It follows that the subsequence $(x_{n_j+1}, w_{n_j+1}, v_{n_j+1})$ also converges to $(x^\dagger, w^\dagger, v^\dagger)$ and that $(x^\dagger, w^\dagger, v^\dagger)$ satisfies the fixed-point equations (9). We can therefore choose $(\hat{x}, \hat{w}, \hat{v}) = (x^\dagger, w^\dagger, v^\dagger)$ in relation (26) to find:

$$\begin{aligned}
& \|U(x^\dagger - x^{N+1})\|^2 + \|V(w^\dagger - w^{N+1})\|^2 + \alpha\|v^\dagger - v^{N+1}\|^2 \\
& \leq \|U(x^\dagger - x^M)\|^2 + \|V(w^\dagger - w^M)\|^2 + \alpha\|v^\dagger - v^M\|^2
\end{aligned}$$

for all $N > M$. As there is a convergent subsequence of (x^n, w^n, v^n) , the right hand side of this expression can be made arbitrarily small by taking $M = n_j$ large enough. Hence the left hand side will be arbitrarily small for all N larger than this M . This proves convergence of the whole sequence (x^n, w^n, v^n) to $(x^\dagger, w^\dagger, v^\dagger)$.

As $(x^\dagger, w^\dagger, v^\dagger)$ satisfies the fixed-point equations, it follows from Theorem 1 that x^\dagger is a solution to problem (1). \square

4 Discussion

- If $\|\frac{1}{2}K^TK + B^TB\| \geq 1$ or $\|AA^T\| \geq 1$ one can rescale the matrices and the variables to arrive at the following iteration:

$$\begin{cases} \bar{v}^{n+1} &= v^n - (Bx^n - b) \\ \bar{x}^{n+1} &= x^n + \tau_1 K^T(y - Kx^n) + \tau_3 B^T \bar{v}^{n+1} - \tau_1 A^T w^n \\ w^{n+1} &= \mathbb{P}_\lambda(w^n + \frac{\tau_2}{\tau_1} A \bar{x}^{n+1}) \\ x^{n+1} &= x^n + \tau_1 K^T(y - Kx^n) + \tau_3 B^T \bar{v}^{n+1} - \tau_1 A^T w^{n+1} \\ v^{n+1} &= v^n - \frac{1}{\alpha}(Bx^{n+1} - b) \end{cases} \quad (27)$$

with step size parameters $\tau_1, \tau_2, \tau_3 > 0$ that satisfy $\|\tau_1 K^T K / 2 + \tau_3 B^T B\| < 1$ and $\tau_2 \|AA^T\| < 1$.

- The ℓ_1 -norm in problem (1) does not necessarily have to be defined as $\|u\|_1 = \sum_i |u_i|$. In section 5 we will use an ℓ_1 -norm of the form $\|u\|_1 = \sum_i \max(|u_{i,1}|, \dots, |u_{i,m}|)$ for a vector $u \in \mathbb{R}^{N \times m}$. Such a penalty is useful for promoting joint sparsity on the $u_{i,j}$ (for a fixed i). Indeed, if e.g. $u_{i,1}$ is non-zero, then all other $u_{i,j}$ ($j \neq 1$) may be as large as $|u_{i,1}|$ as well, without increasing $\max(|u_{i,1}|, \dots, |u_{i,m}|)$. As $\lambda \max(|z_1|, \dots, |z_m|) = \max_{\|w\|_1 \leq \lambda} \langle w, z \rangle$, one needs to replace the projection \mathbb{P}_λ in (10) by N projections on an ℓ_1 -ball (in \mathbb{R}^m) of radius λ . We denote the projection on an ℓ_1 -ball of radius λ in \mathbb{R}^m by Q_λ . In algorithm (15), that is used for the special case $A = \text{Id}$, one has to replace the component-wise soft-thresholding S_λ with a new thresholding function $T_\lambda = \text{Id} - Q_\lambda$ (and apply it to N vectors of size m). The operator T_λ can be computed as follows [13]. Let $z \in \mathbb{R}^m$ and order the entries such that $|z_{i_1}| \geq |z_{i_2}| \geq \dots \geq |z_{i_m}|$. Then:

$$\begin{cases} \text{for } \|z\|_1 \leq \lambda & : T_\lambda(z) = 0 \\ \text{for } \|z\|_1 > \lambda & : \begin{cases} (T_\lambda(z))_{i_j} = \text{sgn}(z_{i_j})(\sum_{k=1}^l |z_{i_k}| - \lambda)/l & j = 1, \dots, l \\ (T_\lambda(z))_{i_j} = z_{i_j} & j = l+1, \dots, m \end{cases} \end{cases} \quad (28)$$

where $l \in \{1, \dots, m\}$ is the largest index satisfying $|z_{i_l}| \geq (\sum_{k=1}^l |z_{i_k}| - \lambda)/l$.

- The ℓ_1 -norm in functional (1) can be replaced by a convex lower semi-continuous function H :

$$\hat{x} = \arg \min_{Bx=b} \|Kx - y\|^2 + 2H(Ax). \quad (29)$$

(assuming a minimizer exists). The projection operator \mathbb{P}_λ in algorithm (10) then needs to be replaced by the proximity operator of the convex

conjugate of H , H^* , defined by $H^*(w) = \sup_x \{\langle w, x \rangle - H(x)\}$ (see e.g. [8]):

$$\begin{cases} \bar{v}^{n+1} &= v^n - (Bx^n - b) \\ \bar{x}^{n+1} &= x^n + K^T(y - Kx^n) + B^T v^{n+1} - A^T w^n \\ w^{n+1} &= \text{prox}_{H^*}(w^n + A\bar{x}^{n+1}) \\ x^{n+1} &= x^n + K^T(y - Kx^n) + B^T \bar{v}^{n+1} - A^T w^{n+1} \\ v^{n+1} &= v^n - \frac{1}{\alpha}(Bx^{n+1} - b), \end{cases} \quad (30)$$

which converges under the same conditions as in Theorem 2 to a minimizer of problem (29). The proximity operator of H is defined as $\text{prox}_H(u) = \arg \min_w H(w) + \|w - u\|^2/2$, and $\text{prox}_{H^*} = \text{Id} - \text{prox}_H$. It is important to remark that only the proximity operator prox_H of H is needed, not the proximity operator of $H(A\cdot)$. In fact the convergence of algorithm (11) (i.e. without the linear constraints $Bx = b$) was proven in this more general context in [17].

- The functional $\mathcal{F}(x) = \|Kx - y\|^2 + 2\lambda\|Ax\|_1$, evaluated in the iterates x^n , does not decrease monotonically as a function of n . Because the iterates x^n do not necessarily satisfy the constraint in every step, it is even possible that $\mathcal{F}(x^n) < \mathcal{F}(\hat{x})$ for some n . The constraint $Bx = b$ is only satisfied in the limit $n \rightarrow \infty$.
- If one wants to solve the ℓ_1 -norm constrained problem

$$\hat{x} = \arg \min_{Bx=b, \|x\|_1 \leq R} \|Kx - y\|^2 \quad (31)$$

instead of the ℓ_1 -norm penalized problem (16), then one may replace the soft-thresholding \mathbb{S}_λ in algorithm (15) by projection on the ℓ_1 -ball. The algorithm is:

$$\begin{cases} \bar{w}^{n+1} &= w^n - (Bx^n - b) \\ x^{n+1} &= Q_R(x^n + K^T(y - Kx^n) + B^T \bar{w}^{n+1}) \\ w^{n+1} &= w^n - \frac{1}{\alpha}(Bx^{n+1} - b), \end{cases} \quad (32)$$

where Q_R is the projection on the ℓ_1 -ball of radius R (such a projection is explicitly doable by computer; see expression (28), with λ replaced by R , and the paragraph above). This algorithm converges for $\|\frac{1}{2}K^T K + B^T B\| < 1$ and $\alpha > 1/2$. This can be shown by using Lemma 1 (for an ℓ_1 -ball instead of an ℓ_∞ -ball) and proceeding in the same way as in Theorem 2 (without proof).

- In the special case when $A = \text{Id}$, the algorithm (15) could also have been obtained from algorithm (A_0), formula (3.2) of [23] (it wasn't done explicitly). This can be achieved by the choice $H \rightarrow \frac{1}{2}\|Kx - y\|^2$, $A \rightarrow B$, $J \rightarrow \|\cdot\|_1$ and $Q_0 \rightarrow \text{Id} - K^T K - B^T B$ and $C \rightarrow \alpha$ in [23]. However, the assumption in [23] that Q_0 is positive definite (necessary to prove the convergence of the algorithm), amounts to having $\|K^T K + B^T B\| < 1$. We have shown here that the condition $\|\frac{1}{2}K^T K + B^T B\| < 1$ is already sufficient to guarantee convergence.

As already mentioned, when $A = \text{Id}$ and $K = 0$ the proposed algorithm (10) reduces to the algorithm (18) for the problem (17). This algorithm was also (re)derived in [23] under a slightly different form (see equation (5.6) of [23]).

5 Application to magneto-encephalography

The goal of magneto-encephalography (MEG) is to determine a current density \vec{J} in the brain by measuring (a component of) the magnetic field \vec{B} induced by \vec{J} , in several points outside the scalp. We assume that \vec{B} and \vec{J} are linked by the Biot-Savart law:

$$\vec{B}(\vec{r}) = \frac{\mu_0}{4\pi} \int_V \vec{J}(\vec{r}') \times \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} dV', \quad (33)$$

with $\mu_0 = 4\pi \times 10^{-7} \text{Vs/Am}$ and V the volume in which the current flows. The conservation of charges implies that $\text{div}(\vec{J}) = 0$. For more information, see [11] and references therein.

In this section, we pose and solve a synthetic inverse problem inspired by this problem. We consider a thin spherical shell V centered at the origin, with an outer radius of 9cm and a thickness of 1mm. Measurements are made at 500 random points \vec{r}_i ($i : 1 \dots 500$) uniformly distributed on the upper hemisphere with a distance of 10cm to the origin. The data is composed of the radial component of the magnetic field $B_r(\vec{r})$ measured in these 500 points. For simplicity all the current densities considered below do not depend on the radius and do not have a radial component. In other words, we treat a 2D problem.

In order to discretize the problem, we use the ‘‘cubed sphere’’ parametrization introduced in [19]; it maps the sphere to the six sides of a cube and a regular grid is then used on each of these six sides. We choose a grid with 64^2 voxels on each side.

As input model (that we will want to reconstruct) we choose the current density distribution \vec{J}_{in} shown in Figure 1. This model is divergence-free

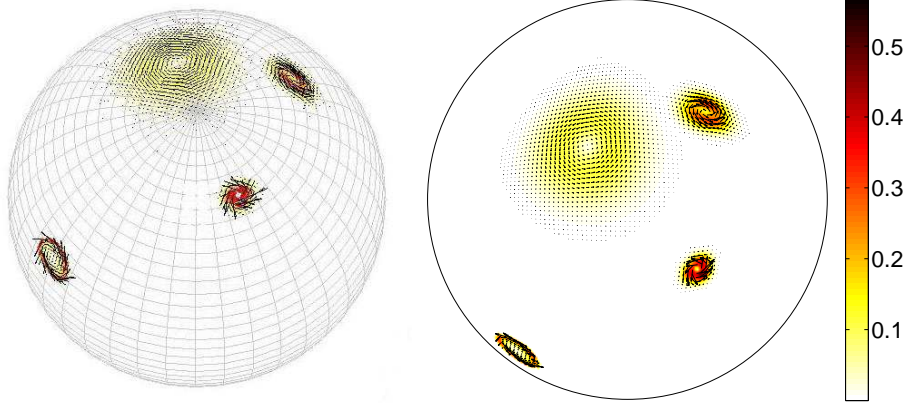


Figure 1: Input model \vec{J}_{in} on the sphere (left) and its top view (right). The background color map is proportional to the local norm of the current density vector \vec{J}_{in} .

by construction. The matrix K encodes the Biot-Savart law for the radial component $B_r(\vec{r}_i) = \vec{B} \cdot \vec{e}_r(\vec{r}_i)$ of \vec{B} in the 500 measurement points \vec{r}_i ($i : 1 \dots 500$). As

$$\begin{aligned} \vec{B}(\vec{r}_i) \cdot \vec{e}_r(\vec{r}_i) &= \frac{\mu_0}{4\pi} \int_V \left(\vec{J}(\vec{r}') \times \frac{\vec{r}_i - \vec{r}'}{|\vec{r}_i - \vec{r}'|^3} \right) \cdot \vec{e}_r(\vec{r}_i) dV' \\ &= \frac{\mu_0}{4\pi} \int_V \left(\frac{\vec{r}_i - \vec{r}'}{|\vec{r}_i - \vec{r}'|^3} \times \vec{e}_r(\vec{r}_i) \right) \cdot \vec{J}(\vec{r}') dV', \end{aligned} \quad (34)$$

we set:

$$(K\vec{J})_i = \sum_{\text{all voxels}} K_{i,\text{voxel}} \cdot \vec{J}_{\text{voxel}}, \quad K_{i,\text{voxel}} = \frac{\mu_0}{4\pi} \int_{\text{voxel}} \frac{\vec{r}_i - \vec{r}'}{|\vec{r}_i - \vec{r}'|^3} \times \vec{e}_r(\vec{r}_i) dV' \quad (35)$$

The data y is obtained from the synthetic input model \vec{J}_{in} . More precisely, the data is constructed by setting $y = K\vec{J}_{\text{in}} + \epsilon$, where ϵ is Gaussian noise. We choose a noise level of 10%: $\|\epsilon\| = 0.1 \times \|K\vec{J}_{\text{in}}\|$.

Our goal now is to reconstruct the current density on each of the voxels of the cubed sphere from the noisy data y . The cubed sphere parametrization allows us to use a simple set of wavelet functions on the sphere introduced in [21]. They belong to the CDF 4-2 family [7]. As the current density has two components, we have to reconstruct $64^2 \times 6 \times 2 = 49152$ coefficients from merely 500 measurements. The problem is therefore severely under-determined.

To impose the assumed sparsity of \vec{J} in the wavelet basis, we will use an ℓ_1 -penalized least squares functional, with the additional linear constraint

$\text{div}(\vec{J}) = 0$. We set \vec{w} equal to the list of coefficients of \vec{J} in the wavelet basis. Each element \vec{w}_k of the list \vec{w} (and \vec{J}) has two components $w_{k,1}, w_{k,2}$ corresponding to the two angular directions on the cubed sphere. The wavelet transform that maps the wavelet coefficients \vec{w} to the current density \vec{J} is represented by the operator W^{-1} so that $\vec{J} = W^{-1}\vec{w}$. In fact W^{-1} works on both angular components separately.

The minimization problem we want to solve is:

$$\vec{w}_{\text{rec}} = \arg \min_{\vec{w}} \|KW^{-1}\vec{w} - y\|^2 + 2\lambda H(\vec{w}), \quad (36)$$

where the reconstructed current density is $\vec{J}_{\text{rec}} = W^{-1}\vec{w}_{\text{rec}}$. We consider 4 distinct cases:

- a. problem (36) with $H(\vec{w}) = \sum_k |w_{k,1}| + |w_{k,2}|$ without the constraint $\text{div}\vec{J} = 0$
- b. problem (36) with $H(\vec{w}) = \sum_k |w_{k,1}| + |w_{k,2}|$ with the constraint $\text{div}\vec{J} = 0$
- c. problem (36) with $H(\vec{w}) = \sum_k \max\{|w_{k,1}|, |w_{k,2}|\}$ without the constraint $\text{div}\vec{J} = 0$
- d. problem (36) with $H(\vec{w}) = \sum_k \max\{|w_{k,1}|, |w_{k,2}|\}$ with the constraint $\text{div}\vec{J} = 0$

(the sum over k is over all voxels). We want to remark that none of the four penalties considered here are rotationally invariant

Another possibility to reconstruct \vec{J} under the constraint $\text{div}\vec{J} = 0$, is to take $\vec{J} = \text{curl}(G\vec{1}_r)$, with G a scalar field and to reconstruct a sparse G . However, even if the reconstructed G has a small reconstruction error, the error on \vec{J} can be much larger as a result of taking the curl.

Problems (a) and (b) can be solved using algorithm (15) with S_λ given componentwise in expression (2). As there is no constraint involved in problem (a), (a) can be solved more efficiently with the accelerated soft-thresholding algorithm FISTA [1]. We use 2000 iterations. The problem (b) is solved with 20000 iterations of algorithm (15).

For method (c) and (d), joint sparsity is assumed in the wavelet basis. We therefore simply replace the function S_λ used in the algorithms for problems (a) and (b) by the nonlinear operator T_λ defined in formula (28). In other words, we use respectively 2000 iterations of the FISTA algorithm for problem (c) and 20000 iterations of algorithm (15) for problem (d).

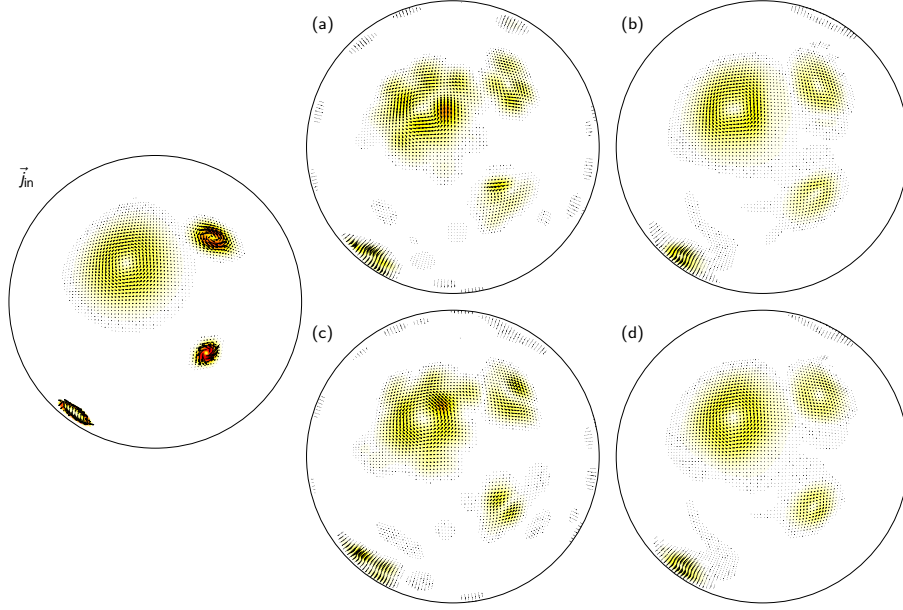


Figure 2: Input model \vec{J}_{in} and four different reconstructions \vec{J}_{rec} . Each current density reconstruction is the minimizer of its own functional. They are numbered according to the list in Section 5. The color scale is the same as in Figure 1.

In each of the four methods, the penalty parameter λ is chosen to fit the data to the noise level ($\|K\vec{J}_{\text{rec}} - y\| = \|\epsilon\|$). The results are displayed in Figure 2. It should be emphasized that the four reconstructions minimize different functionals (with or without additional constraint). The four reconstructions are therefore not identical, not even in the limit of infinitely many iterations (four different limits).

When comparing the results, we first conclude that the FISTA algorithm converges faster to its fixed-point than that algorithm (15) converges to its fixed-point. This is shown in Figure 3 for cases (a) and (b). This is to be expected as FISTA is an accelerated algorithm and (15) is not.

We also calculate the relative reconstruction error $e_{\text{rec}} = \|\vec{J}_{\text{in}} - \vec{J}_{\text{rec}}\| / \|\vec{J}_{\text{in}}\|$, in the four cases (a)–(d). The results are displayed in Table 1. The methods that take the constraint into account perform somewhat better than the ones that do not take the constraint into account. Although the difference is quite small, it does not seem to be a fluctuation: the same result was obtained for other input models, in the same wavelet basis as well as in other bases.

The values of $\|\text{div}\vec{J}\|$ for the four reconstructions are also reported in Table 1. As expected, the constraint is far better satisfied when using meth-

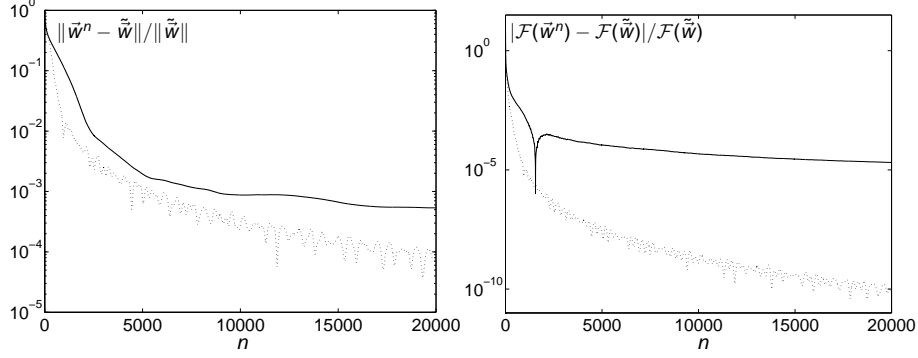


Figure 3: Convergence speed of algorithm (15) (solid) and of FISTA (dotted) to their respective limits as a function of the number of iterations for the problems (a) and (b). Left: Relative distances to the respective limits of the algorithms. Right: Difference of the functionals to the respective final values. In both cases the fixed point \vec{w} of the iterations was obtained using 500000 iterations in their respective algorithms. In case of constrained minimization, the functional value may go below the final limit value. This is due to the fact that the iterates do not satisfy the divergence-free constraint at every step. Again, the two algorithms solve a different minimization problem.

	a	b	c	d
e_{rec}	0.81	0.75	0.78	0.71
$\ \text{div} \vec{J}_{\text{rec}}\ $	0.59	$3.9 \cdot 10^{-6}$	0.61	$8.3 \cdot 10^{-6}$
$\#$ nonzero in \vec{w}_{rec}	82	9780	194	18108

Table 1: Reconstruction data for the four simulations discussed in Section 5.

ods (b) and (d) with algorithm (15), than methods (a) and (c). One of the consequences is that the methods (b) and (d) give better visual results, as can be observed in Figure 2. While all four methods localize the objects in \vec{J} quite well, only those corresponding to divergence-free reconstructions are well structured.

Finally, the divergence-free reconstructions are less sparse than the reconstructions that do not satisfy this constraint. For example when solving case (a) with the FISTA algorithm, only 82 of the 49152 coefficients (in the wavelet basis) are nonzero, while case (b) solved with algorithm (15), gives 9780 nonzero coefficients. This is due to the high number of linear constraints (one for every voxel).

6 Acknowledgements

I.L. is research associate of the Fonds de la recherche Scientifique-FNRS (Belgium). Part of this research was done while the authors were at the Computational and Applied Mathematics Programme of the Vrije Universiteit Brussel and was supported by VUB GOA-062 and by the Fonds voor Wetenschappelijk Onderzoek-Vlaanderen grant G.0564.09N. The authors would like to thank M. Fornasier and F. Pitolli for the useful discussions on MEG and the referees for their constructive comments.

References

- [1] A. Beck and M. Teboulle. A fast iterative shrinkage-threshold algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.
- [2] S. Bonettini and V. Ruggiero. An alternating extragradient method for total variation-based image restoration from Poisson data. *Inverse Problems*, 27(9):095001, 2011.
- [3] J. Brodie, I. Daubechies, C. De Mol, D. Giannone, and I. Loris. Sparse and stable Markowitz portfolios. *Proc. Natl. Acad. Sci. USA*, 106(30):12267–12272, 2009.
- [4] A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev.*, 51(1):34–81, 2009.
- [5] E. Candès and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Proc. Mag.*, 25(2):21–30, 2008.

- [6] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, 1998.
- [7] A. Cohen, I. Daubechies, and J. Feaveau. Biorthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 45(5):485–560, 1992.
- [8] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer-Verlag, 2011.
- [9] I. Daubechies. *Ten lectures on wavelets*. SIAM, 1992.
- [10] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57(11):1413–1457, 2004.
- [11] C. Del Gratta, V. Pizzella, F. Tecchio, and G. Romani. Magnetoencephalography - a noninvasive brain imaging method with 1 ms time resolution. *Rep. Prog. Phys.*, 64(12):1759–1814, 2001.
- [12] M. Fornasier and F. Pitolli. Adaptive iterative thresholding algorithms for magnetoencephalography (MEG). *J. Comput. Appl. Math.*, 221(2):386–395, 2008.
- [13] M. Fornasier and H. Rauhut. Recovery algorithms for vector-valued data with joint sparsity constraints. *SIAM J. Numer. Anal.*, 46(2):577–613, 2008.
- [14] P. G. Lemarie-Rieusset. Ondelettes vecteurs à divergence nulle. *C. R. Acad. Sc. Paris*, 313(5):213–216, 1991.
- [15] P. G. Lemarie-Rieusset. Analyses multi-résolutions non orthogonales, commutation entre projecteurs et dérivation et ondelettes vecteurs à divergence nulle. *Rev. Mat. Iberoamericana*, 8(2):221–237, 1992.
- [16] W. G. Litvinov, T. Rahman, and X.-C. Tai. A modified TV-Stokes model for image processing. *SIAM J. Sci. Comput.*, 33(4):1574–1597, 2011.
- [17] I. Loris and C. Verhoeven. On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty. *Inverse Problems*, 27(12):125007, 2011.

- [18] S. Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, third edition, 2009.
- [19] C. Ronchi, R. Iacono, and P. Paolucci. The “cubed sphere”: A new method for the solution of partial differential equations in spherical geometry. *J. Comput. Phys.*, 124(1):93–114, 1996.
- [20] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, 1992.
- [21] F. J. Simons, I. Loris, G. Nolet, I. C. Daubechies, S. Voronin, J. S. Judd, P. A. Vetter, J. Charléty, and C. Vonesch. Solving or resolving global tomographic models with spherical wavelets, and the scale and sparsity of seismic heterogeneity. *Geophys. J. Int.*, 187(2):969–988, 2011.
- [22] R. Stevenson. Divergence-free wavelet bases on the hypercube: free-slip boundary conditions, and applications for solving the instationary Stokes equations. *Math. Comp.*, 80(275):1499–1523, 2011.
- [23] X. Zhang, M. Burger, and S. Osher. A unified primal-dual algorithm framework based on Bregman iteration. *J. Sci. Comput.*, 46(1):20–46, 2011.